Discussant Remarks

Kunal Talwar

CSAC Meeting Dec 6-7, 2018

Outline

- Why modernizing DA is necessary?
- What is Differential Privacy?
- Questions for the Census Bureau

Fundamental Law of Information Recovery

“If you release too many statistics about a dataset, it can be reconstructed”

Fundamental Law of Information Recovery

“If you release too many statistics about a dataset, it can be reconstructed”

True even if the statistics are released with limited noise.

True even if some of the statistics have unbounded noise.

Re-identification

When data is high dimensional, pseudo-identifiers exist

Re-identification

When data is high dimensional, pseudo-identifiers exist

Sweeney easily found Gov. Weld’s medical records


NETFLIX
Netflix Prize
Home
Rules
Leaderboard
Register
Update
Submit
Downlo
NETFLIX
Browse
Recommendations
Friends
Queue
Buy DVDs
Home
Genres
New Releases
Previews
Netflix Top 100
Movies For You
Randy, the following movies were chosen based on your interest in:
Bowling for Columbine
Carnivale: Season 1
Fahrenheit 9/11
The Big One
All Discs Guaranteed!
You really liked it...
Now
for just $5.99
Shop
titles
as low
Original art
Kunal Talwar

"An adversary who knows a little bit about some subscriber can easily identify her record if it is present in the dataset, or, at the very least, identify a small set of records which include the subscriber's record"

-Narayanan and Shmatikov

Disclosure Avoidance at the Census

Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing

Laura McKenna[1]

October 2018

One takeaway: The Census Bureau's disclosure avoidance techniques have evolved over the years. They have always tried to use the state-of-the-art techniques.

Disclosure Avoidance at the Census

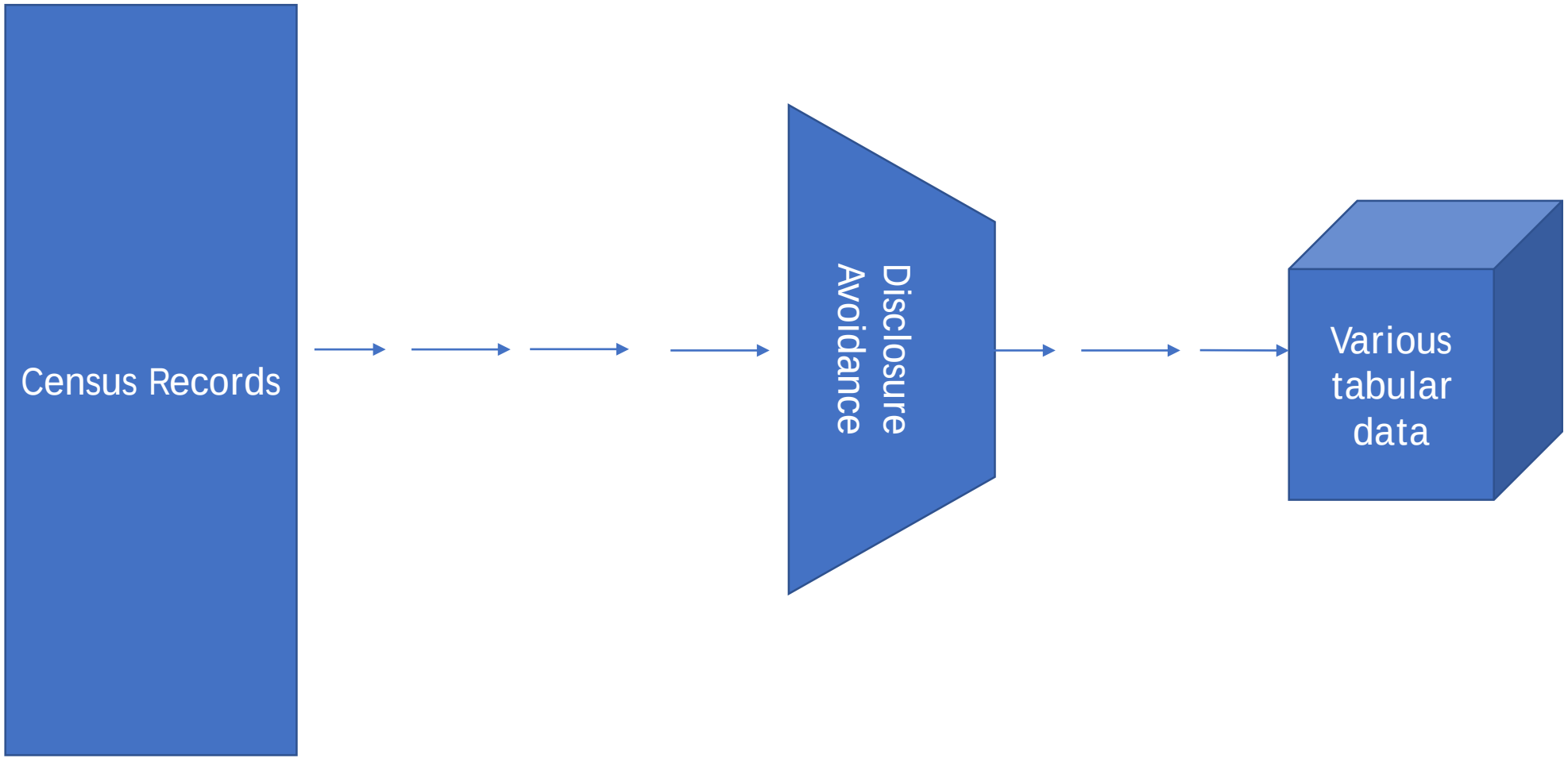
Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing

Laura McKenna[1]

October 2018

One takeaway: The Census Bureau's disclosure avoidance techniques have evolved over the years. They have always tried to use the state-of-the-art techniques.

Currently accepted State-of-the-art: Differential Privacy


Census Records
Disclosure Avoidance
Various tabular data

Differential Privacy

Algorithm: Randomized function M : Dataset $\rightarrow$ Tables

Dataset D and D' are neighbors if they differ in one person's data.

DP : $M(D) \approx M(D')$. Privacy Parameter ϵ measures distance between distributions (worst case over all pairs D and D')

Properties

- Post-Processing : If M is ϵ-DP, then for any f, $f\big(M(D)\big)$ is ϵ-DP

- Composition : If M_1 and M_2 are ϵ-DP, then $M_1; M_2$ is 2ϵ-DP

Allow us to reason about multiple releases

Allow us to build complex algorithms out of simple building blocks

Differential Privacy

Rich research literature on design and analysis of differentially private algorithms

Used in Google Chrome, Apple iOS, Microsoft Windows, Snapchat, Uber.

Preventing Reconstruction

- Semantic Properties of DP imply that reconstruction is not possible

- Answers are noisy
- Noise grows with the number of queries answered

Preventing Reconstruction

- Semantic Properties of DP imply that reconstruction is not possible

- Answers are noisy
- Noise grows with the number of queries answered

- For count queries:

Error of DP algorithm $\approx$ Error needed to prevent reconstruction

Transparency

- Kerckhoffs's law of cryptography:

  When designing secure systems, assume that the adversary knows everything except for the secret key

Bonus: analyst can account for the noise added by mechanism.

Questions

Q1: CDP/RDP parameters for the implemented algorithms may be a lot better. Why not also publish those numbers?

Q2: There are various trade-offs within the algorithm. Are the benchmarks to help choose among those?

Q3: Publicizing the use of DP (and the parameters)?

Questions and Suggestions

Q4: Are the examples of specific analyses of the kind consumers of this data do, that you plan to benchmark?

Q5: Invariants seem to hurt the privacy guarantee. Which invariants can be relaxed?